A NONLINEAR MODEL FOR ESTIMATING PROBABILITIES OF k EVENTS [1]

RICHARD H. JONES
Department of Mathematics, University of Hawaii [2], Honolulu, Hawaii

ABSTRACT

Equations are derived for the estimation of the parameters in a nonlinear model for the probability of more than two mutually exclusive and exhaustive events. The estimated probabilities are between zero and one and sum to one. The equations for least squares and maximum likelihood estimation are given, and it is pointed out that the maximum likelihood estimate has the form of weighted least squares with estimated weights giving more weight to low probability events.

In a recent paper by Brelsford and Jones [1], the fitting of a logistic curve to zero-one data using least squares and maximum likelihood was presented using a technique developed by Walker and Duncan [2]. The paper stated that the dichotomous case generalizes to a model suggested by Cox [3] for estimating probabilities for more than two events. Several requests have been received for details of this generalization, and this note provides the details together with more explicit derivations.

Suppose there are k mutually exclusive and exhaustive events. For the observation at time t let $z_i(t)=1$ if event i occurs and zero otherwise. This gives

$$\sum_{i=1}^{k} z_i(t)=1. \tag{1}$$

The model to be fitted to the data is

$$P_i(t)=\text{Prob}\,\{z_i(t)=1\}=\frac{\exp\left[\sum_{j=1}^{p} x_j(t)\beta_{ij}\right]}{\sum_{l=1}^{k}\exp\left[\sum_{j=1}^{p} x_j(t)\beta_{lj}\right]} \tag{2}$$

where β_{ij} are the regression coefficients to be estimated and $x_j(t)$ are the predictor variables. This model has the properties

$$0<P_i(t)<1 \tag{3}$$

$$\sum_{i=1}^{k} P_i(t)=1. \tag{4}$$

Because of the constraint (4), without loss of generality we may put

$$\beta_{ij}=0 \text{ for } i=1 \tag{5}$$

which gives $p(k-1)$ regression coefficients to be estimated. Since there are $k-1$ linearly independent $z_i(t)$ at each time point, the regression equation can be written

$$z_i(t)=P_i(t)+\epsilon_i(t), \qquad i=2, k. \tag{6}$$

The regression coefficients occur nonlinearly so some form of iterative solution is necessary. The equations can be linearized by expanding $P_i(t)$ in a series about an initial guess at the regression coefficients, β_{ij}^0,

$$P_i(t)=P_i^0(t)+\sum_{m=2}^{k}\sum_{n=1}^{p}\frac{\partial}{\partial\beta_{mn}}P_i(t)]_0(\beta_{mn}-\beta_{mn}^0)+\ldots. \tag{7}$$

But

$$\frac{\partial}{\partial\beta_{mn}}P_i(t)=x_n(t)P_i(t)[\delta_{im}-P_m(t)] \tag{8}$$

where

$$\delta_{im}=\begin{cases}1 \text{ if } i=m\\ 0 \text{ otherwise,}\end{cases}$$

so the linearized regression equation is

$$z_i(t)-P_i^0(t)\simeq\sum_{m=2}^{k}\sum_{n=1}^{p} x_n(t)P_i^0(t) \times[\delta_{im}-P_m(t)](\beta_{mn}-\beta_{mn}^0)+\epsilon_i(t). \tag{9}$$

Given observations of $z_i(t)$ and $x_n(t)$, this equation can be solved for $\beta_{mn}-\beta_{mn}^0$, giving a correction to the initial guess β_{mn}^0. This correction is added to the initial guess and the result used as the new guess for the next iteration. The process is repeated until convergence. If the iterations diverge, it indicates a bad first guess and a new attempt can be made. Solving the equations by least squares will give estimates which are best when used against a squared error scoring system.

Maximum likelihood estimates of the regression coefficients can also be obtained. Using the multinomial distribution

$$P_1(t)^{z_1(t)}P_2(t)^{z_2(t)}\ldots P_k(t)^{z_k(t)}, \tag{10}$$

one obtains the likelihood for independent samples

$$L=\prod_t\prod_{l=1}^{k} P_l(t)^{z_l(t)}. \tag{11}$$

The log likelihood is

$$\log L=\sum_t\sum_{l=1}^{k} z_l(t)\log P_l(t). \tag{12}$$

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR Contract No. AF49(638)-1302.
[2] On leave from the Johns Hopkins University.

Differentiating with respect to β_{ij} and equating the result to zero gives

$$\sum_t \sum_{l=1}^{k} z_l(t)x_j(t)[\delta_{li}-P_i(t)]=0. \quad (13)$$

Noting that

$$z_i(t)=1-\sum_{l\neq i} z_l(t),$$

one can write equation (13) as

$$-\sum_t \sum_{l\neq i} z_l(t)x_j(t)P_i(t)+\sum_t \left[1-\sum_{l\neq i} z_l(t)\right]x_j(t)[1-P_i(t)]=0. \quad (14)$$

This reduces to

$$\sum_t z_i(t)x_j(t)=\sum_t P_i(t)x_j(t) \quad (15)$$

giving the system of $p(k-1)$ equations which must be solved for β_{ij}. Again the right hand side can be linearized by expanding $P_i(t)$ in a series giving

$$\sum_t [z_i(t)-P_i^0(t)]x_j(t)\simeq \sum_t x_j(t) \sum_{m=2}^{k} \sum_{n=1}^{p} x_n(t)P_i^0(t)[\delta_{im}-P_m(t)](\beta_{mn}-\beta_{mn}^0)$$
$$i=2,k;\ j=1,p. \quad (16)$$

The solution of these equations requires the inversion of a $p(k-1)$ by $p(k-1)$ matrix at each iteration. Equation (9) is a linearized regression equation with errors, $\epsilon_i(t)$, which do not have constant variance. A weighted least squares analysis would be indicated; however, the variances and therefore the weights involve the regression coefficients which are unknown. An unweighted least squares solution can be obtained by iteration. The maximum likelihood equations have the form of a weighted least squares solution except that the weights involve the estimated regression coefficients. Therefore, more weight is given to low probability events. Equation (16) is simpler than the normal equations calculated from (9), and the solution minimizes the logarithmic score.

The method presented by Walker and Duncan [2] for estimating the probability of k events may be preferable when the events have a natural ordering since it involves the estimation of fewer regression coefficients.

There is a rapidly growing literature on nonlinear regression, and a clear account together with a long list of references is given in Draper and Smith [4]. The method discussed in this paper is the usual Gauss-Newton procedure. A difficulty which is sometimes encountered is iterations which oscillate and converge slowly. It is usually possible to damp these oscillations and speed convergence by adding a positive constant to the diagonal elements of the matrix before it is inverted. This gives a combination of the Gauss-Newton method and the method of steepest descent.

REFERENCES

1. W. M. Brelsford and R. H. Jones, "Estimating Probabilities," *Monthly Weather Review*, vol. 95, No. 8, Aug. 1967, pp. 570–576.
2. S. H. Walker and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several Independent Variables," *Biometrika*, London, vol. 54, No. 1 and 2, June 1967, pp. 167–179.
3. D. R. Cox, "Some Procedures Connected with the Logistic Qualitative Response Curve," *Research Papers in Statistics* (F. N. David, Ed.), John Wiley and Sons, London, 1966, pp. 55–71.
4. N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley and Sons, New York, 1966, 407 pp.

[*Received November 27, 1967; revised January 12, 1968*]